

Spatio-Temporal Smoothing of CO₂ Retrievals

Noel Cressie

Program in Spatial Statistics and Environmental Statistics

Department of Statistics, The Ohio State University

Matthias Katzfuss

Institut für Angewandte Mathematik, Universität Heidelberg

Acknowledgment: AIST grant from NASA's ESTO (PI: Amy Braverman)

AIRS CO_2 Data

Global satellite measurements of CO_2 from the AIRS instrument (data below $60^\circ S$ are not available)

Challenges of global remote-sensing data:

- **Massiveness**

- Need dimension reduction

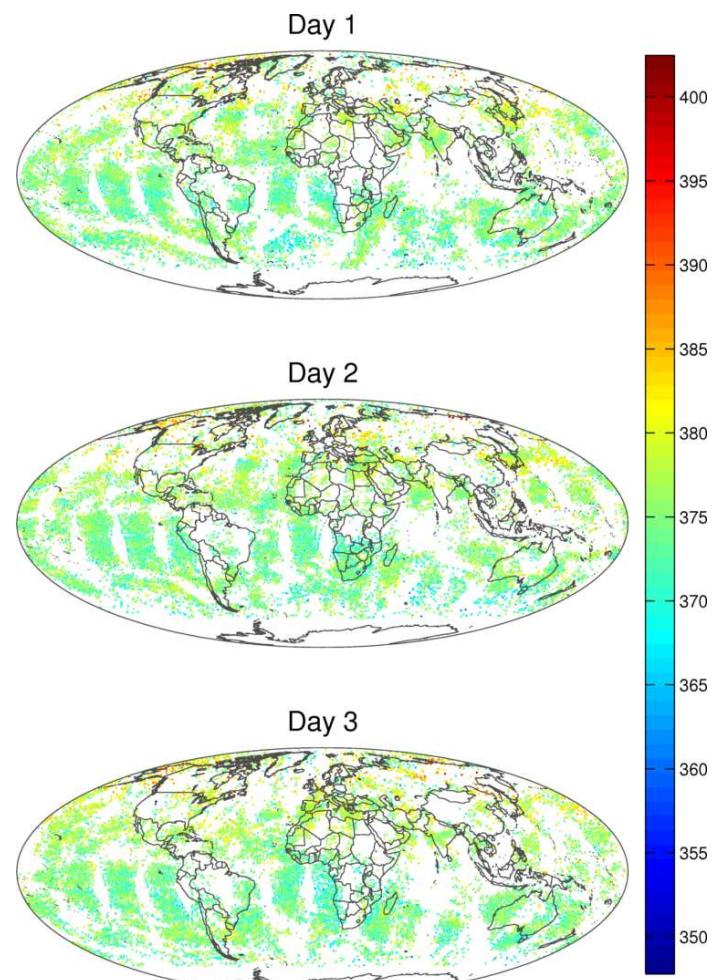
- **Sparseness**

- Need to take advantage of spatial and temporal correlations

- **Nonstationarity**

- Need a flexible model

Goal: Using AIRS CO_2 data, **map CO_2** globally after “de-noising” and “gap-filling”; also **map uncertainty**



Hierarchical Spatio-Temporal Model

- Time is **discrete**
- Write the **hidden** spatio-temporal process as $Y_t(\mathbf{s})$ at time t and location \mathbf{s}
- The true process is observed **imperfectly** and **incompletely**
- **Data Model:** $Z_t(\mathbf{s}) = Y_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$
 - $\varepsilon_t(\cdot) \sim Gau(0, \sigma_{\varepsilon,t}^2 v_\varepsilon(\cdot))$, is measurement error that is assumed uncorrelated across time and space
 - $\sigma_{\varepsilon,t}^2$ and $v_\varepsilon(\cdot)$ are known
- **Goal:** For any \mathbf{s}_0 on the globe, estimate $Y_t(\mathbf{s}_0)$; $t \in \{1, \dots, T\}$, after observing $\{Z_t(\mathbf{s}_{i,t})\}$; define

$$\mathbf{Z}_{1:T} \equiv (\mathbf{Z}'_1, \dots, \mathbf{Z}'_T)', \text{ where } \mathbf{Z}_t \equiv (Z(\mathbf{s}_{1,t}), \dots, Z(\mathbf{s}_{n_t,t}))'$$

- Consider a **time series of spatial processes** (e.g., Cressie and Wikle, 2011). Using a state space model, we **smooth** the data (cf. **filtering**)

Hierarchical Spatio-Temporal Model, ctd.

- **Process Model:** Do not use a transport model. Take a secular approach and use a dynamical statistical model; let the data speak. Model the spatio-temporal variability statistically:

$$Y_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})' \boldsymbol{\beta}_t + \mathbf{S}_t(\mathbf{s})' \boldsymbol{\eta}_t + \xi_t(\mathbf{s})$$

The dimension of $\boldsymbol{\eta}_t$ is $r \ll n_t$ (dimension reduction; e.g., $n_t = 12515$, $r = 380$).

Hierarchical Spatio-Temporal Model, ctd.

- In **vector notation**,

$$\mathbf{Z}_t = \mathbf{Y}_t + \boldsymbol{\varepsilon}_t$$

$$\mathbf{Y}_t = X_t \boldsymbol{\beta}_t + S_t \boldsymbol{\eta}_t + \boldsymbol{\xi}_t$$

$$\Sigma_t \equiv \text{var}(\mathbf{Z}_t) = S_t K_t S_t' + D_t, \text{ where } K_t \equiv \text{var}(\boldsymbol{\eta}_t)$$

$D_t \equiv \sigma_{\xi,t}^2 V_{\xi,t} + \sigma_{\varepsilon,t}^2 V_{\varepsilon,t}$, is a diagonal $n_t \times n_t$ matrix

- **Temporal dynamics** on r -dimensional space

$$\boldsymbol{\eta}_t = H \boldsymbol{\eta}_{t-1} + \boldsymbol{\delta}_t$$

$$U = \text{var}(\boldsymbol{\delta}_t)$$

$$K_t = HK_{t+1}H' + U$$

- The part of the model for \mathbf{Y}_t given by $S_t \boldsymbol{\eta}_t + \boldsymbol{\xi}_t$, is called a **Spatio-Temporal Random Effects (STRE) model**
- **Goal:** Estimate \mathbf{Y}_t from $\mathbf{Z}_{1:T}$

Fixed Rank Smoothing (FRS)

- For the moment, assume **parameters θ are known**

- An extension of the **Kalman smoother** gives, for $t \in \{1, \dots, T\}$,

$$\boldsymbol{\eta}_{t|T} \equiv E(\boldsymbol{\eta}_t | \mathbf{Z}_{1:T}), \quad P_{t|T} \equiv \text{var}(\boldsymbol{\eta}_t | \mathbf{Z}_{1:T}), \quad P_{t,t-1|T} \equiv \text{cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1} | \mathbf{Z}_{1:T})$$

$$\xi_{t|T}(\mathbf{s}_0) \equiv E(\xi_t(\mathbf{s}_0) | \mathbf{Z}_{1:T}), \quad c_{t|T}(\mathbf{s}_0) \equiv \text{var}(\xi_t(\mathbf{s}_0) | \mathbf{Z}_{1:T}), \quad \mathbf{R}_{t|T}(\mathbf{s}_0) \equiv \text{cov}(\boldsymbol{\eta}_t, \xi_t(\mathbf{s}_0) | \mathbf{Z}_{1:T})$$

- Then, **FRS** yields (Cressie, Shi, & Kang, 2010) the estimate of $Y_t(\mathbf{s}_0)$. For $t = 1, \dots, T$:

$$\hat{Y}_t(\mathbf{s}_0) = \mathbf{x}_t(\mathbf{s}_0)' \boldsymbol{\beta}_t + \mathbf{S}(\mathbf{s}_0)' \boldsymbol{\eta}_{t|T} + \xi_{t|T}(\mathbf{s}_0),$$

and the mean squared prediction error (MSPE), $E(\hat{Y}_t(\mathbf{s}_0) - Y_t(\mathbf{s}_0))^2$, can also be calculated. Our **measure of uncertainty is $(MSPE)^{1/2}$**

- Rapid computation:** Need to invert the very large $n_t \times n_t$ covariance matrix Σ_t ; only inversion of $r \times r$ matrices and $n_t \times n_t$ diagonal matrices are required. Use the **Sherman-Morrison-Woodbury identity**, over and over:

$$(I + P K P')^{-1} = I - P(K^{-1} + P' P)^{-1} P'$$

Parameters

- The parameters, θ , consist of $\{\beta_t\}$, $\{\sigma_{\xi,t}^2\}$, and the elements of K_0 , H , and U
- In FRS, θ was assumed known. In this presentation, we take an **empirical-Bayes approach** and “plug in” estimates of the components of θ . Ideally, estimate θ by **maximum likelihood estimation (MLE)**; see Katzfuss and Cressie (2011a)
- A **fully Bayes approach** has also been developed (Katzfuss and Cressie, 2011b)

Empirical Bayes: Estimate the Parameters

- Define **innovations**, $\alpha_t \equiv \mathbf{Z}_t - X_t\beta_t - S_t E(\eta_t | \mathbf{Z}_{1:t-1})$; $t = 1, \dots, T$

$$\alpha_t \stackrel{\text{ind.}}{\sim} Gau(\mathbf{0}, \Sigma_{\alpha_t}), \quad \text{where}$$

$$\Sigma_{\alpha_t} \equiv S_t \text{var}(\eta_t | \mathbf{Z}_{1:t-1}) S_t' + D_t; \quad t = 1, \dots, T$$

- Likelihood** (Shumway & Stoffer, 2006):

$$-2 \log L(\theta) \equiv -2f(\alpha_1, \dots, \alpha_T | \theta)$$

$$= \sum_{t=1}^T \log |\Sigma_{\alpha_t}(\theta)| + \sum_{t=1}^T \alpha_t(\theta)' \Sigma_{\alpha_t}(\theta)^{-1} \alpha_t(\theta)$$

- Finding MLEs analytically is intractable**, even for $T=1$ (Katzfuss and Cressie, 2009)

EM Algorithm: Review

- Suppose we observe $X_{obs} \sim f(x_{obs}|\theta)$, and we are interested in finding the **MLE of θ**
- If maximizing $L(\theta) \equiv f(x_{obs}|\theta)$ is hard, but there are missing data x_{mis} such that maximizing the **complete likelihood**, $L_c(\theta) \equiv f(x_{obs}, x_{mis}|\theta)$, is easy, we can use the **EM algorithm** to find the MLE of θ

The EM Algorithm (Dempster, Laird, and Rubin, 1977): Starting with $\theta^{[0]} = \theta_0$, repeat the following two steps until convergence:

E-Step: Calculate $Q(\theta; \theta^{[l]}) \equiv E_{\theta^{[l]}} \{ \log f(x_{obs}, X_{mis}|\theta) | x_{obs} \}$

M-Step: Calculate $\theta^{[l+1]} = \arg \max_{\theta} Q(\theta; \theta^{[l]})$

The EM Algorithm in this Context (Katzfuss & Cressie, 2011a)

- Reminder: $\mathbf{Z}_t = X_t \boldsymbol{\beta}_t + S_t \boldsymbol{\eta}_t + \boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t$
- Missing (mis) “data”:** $\{\boldsymbol{\eta}_t : t = 0, 1, \dots, T\}$ and $\{\boldsymbol{\xi}_t : t = 1, \dots, T\}$

$$\begin{aligned}-2 \log L_c(\boldsymbol{\theta}) &= -2 \log f(\boldsymbol{\eta}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T} | \boldsymbol{\theta}) \\&= \log |K_0| + \text{tr}(K_0^{-1} \boldsymbol{\eta}_0 \boldsymbol{\eta}'_0) \\&\quad + \sum_{t=1}^T n_t \log \sigma_{\xi,t}^2 + \frac{1}{\sigma_{\xi,t}^2} \sum_{t=1}^T \text{tr}(V_{\xi,t}^{-1} \boldsymbol{\xi}_t \boldsymbol{\xi}'_t) \\&\quad + \sum_{t=1}^T \log |U| + \sum_{t=1}^T \text{tr}\{U^{-1}(\boldsymbol{\eta}_t - H\boldsymbol{\eta}_{t-1})(\boldsymbol{\eta}_t - H\boldsymbol{\eta}_{t-1})'\} \\&\quad + \sum_{t=1}^T \frac{1}{\sigma_{\varepsilon,t}^2} \text{tr}\{V_{\varepsilon,t}^{-1}(\mathbf{Z}_t - X_t \boldsymbol{\beta}_t - S_t \boldsymbol{\eta}_t - \boldsymbol{\xi}_t)(\mathbf{Z}_t - X_t \boldsymbol{\beta}_t - S_t \boldsymbol{\eta}_t - \boldsymbol{\xi}_t)'\}\end{aligned}$$

- E-Step:** FRS provides conditional expectations and covariances of $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ at $\boldsymbol{\theta}^{[l]}$, given $\mathbf{Z}_{1:T}$
- M-Step:** Taking partial derivatives of $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[l]})$ is easy due to terms being additive

Our EM Algorithm

- Choose initial value $\theta^{[0]}$ in the parameter space Θ
- While $\|\theta^{[l+1]} - \theta^{[l]}\| > \delta$:
 1. Run FRS with $\theta^{[l]}$ to obtain $\eta_{t|T}^{[l]}$, $P_{t|T}^{[l]}$, $P_{t,t-1|T}^{[l]}$, $\xi_{t|T}^{[l]}$, and $C_{t|T}^{[l]} \equiv \text{diag}\{c_{t|T}(\mathbf{s}_1), \dots, c_{t|T}(\mathbf{s}_n)\}$
 2. Calculate $\theta^{[l+1]}$ as

$$\beta_t^{[l+1]} = (X_t' V_{\varepsilon,t}^{-1} X_t)^{-1} X_t' V_{\varepsilon,t}^{-1} [\mathbf{Z}_t - S_t \eta_{t|T}^{[l]} - \xi_{t|T}^{[l]}]$$

$$\sigma_{\xi,t}^{2[l+1]} = \text{tr}(V_{\xi,t}^{-1} [C_{t|T}^{[l]} + \xi_{t|T}^{[l]} \xi_{t|T}^{[l]'}]) / n_t$$

$$K_t^{[l+1]} = P_{t|T}^{[l]} + \eta_{t|T}^{[l]} \eta_{t|T}^{[l]}'$$

$$L_t^{[l+1]} = P_{t,t-1|T}^{[l]} + \eta_{t|T}^{[l]} \eta_{t-1|T}^{[l]}'$$

$$H^{[l+1]} = (\sum_{t=1}^T L_t^{[l+1]}) (\sum_{t=0}^{T-1} K_t^{[l+1]})^{-1}$$

$$U^{[l+1]} = (\sum_{t=1}^T K_t^{[l+1]} - H^{[l+1]} \sum_{t=1}^T L_t^{[l+1]'}) / (T - 1)$$

3. Repeat

Properties of the EM Estimators, $\hat{\theta}_{EM}$

- The first $\theta^{[l]}$ that meets the convergence criterion is called $\hat{\theta}_{EM}$
- If $\theta^{[0]} \in \Theta$, then $\theta^{[l]} \in \Theta$, for all $l = 1, 2, \dots$
- Because $Q(\theta; \theta^{[l]})$ is continuous in both arguments, $\hat{\theta}_{EM}$ is generally a solution to the likelihood equations (Wu, 1983)
- For T=1, this algorithm is equivalent to the SRE (spatial-only) EM algorithm (Katzfuss and Cressie, 2009) used in Fixed Rank Kriging (FRK)
- The **computational cost** of the algorithm is **linear in each n_t**

Summary: EM-FRS

- Run EM on the observed data to obtain EM parameter estimates, $\hat{\theta}_{EM}$
- Do FRS with EM estimates plugged in to obtain $\hat{Y}_t(s_0)$ and its corresponding root mean squared prediction error at each location s_0 ; for example,

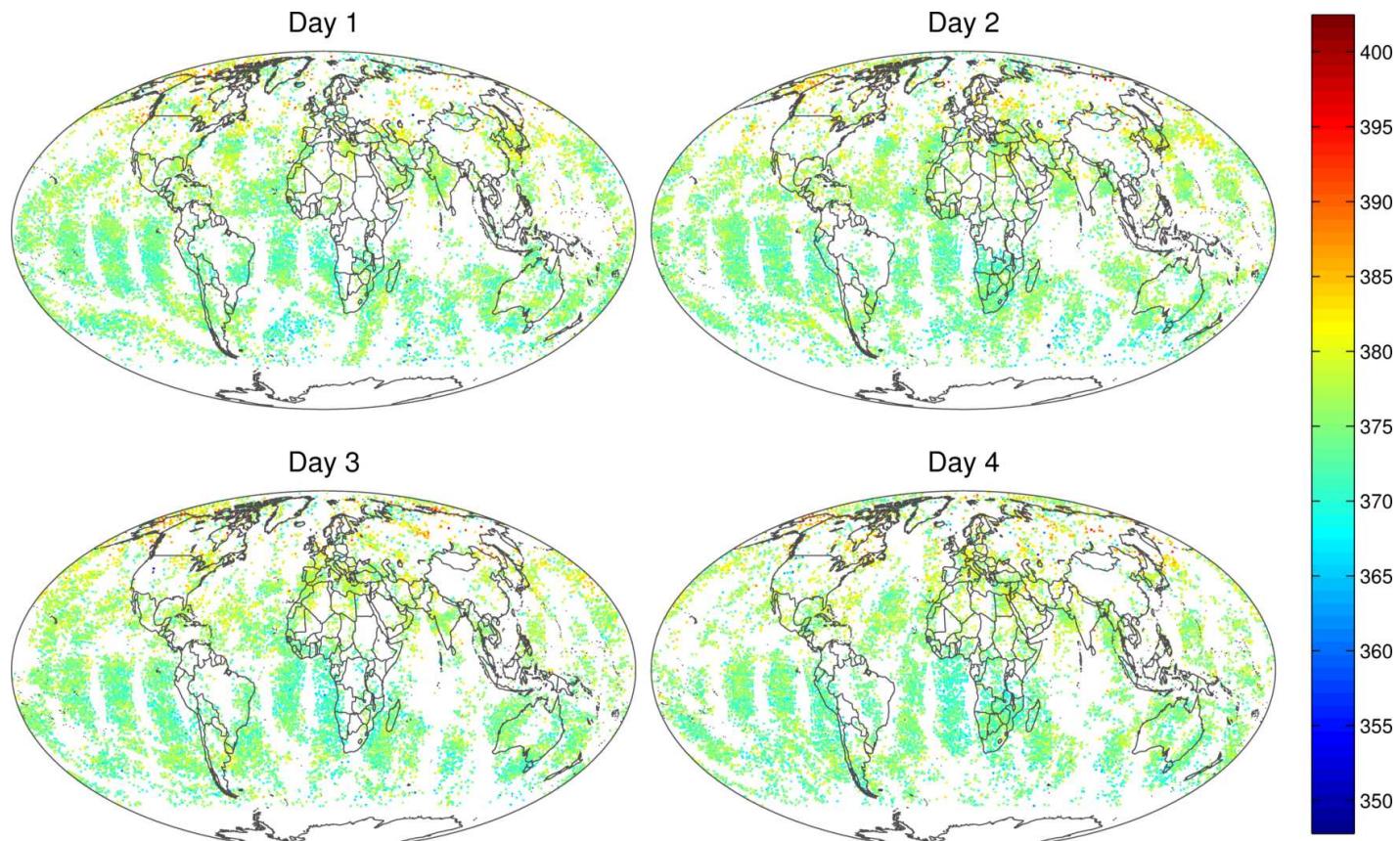
$$\hat{Y}_t(s_0)^{EM} \equiv \mathbf{x}_t(s_0)' \hat{\boldsymbol{\beta}}^{EM} + S(s_0)' \mathbf{n}_{t|T}^{EM} + \xi_{t|T}(s_0)^{EM}$$

AIRS CO_2 Data

- Measurements are taken by the Atmospheric Infrared Sounder (**AIRS**) instrument, on NASA's Aqua satellite
- **Mid-tropospheric CO_2** at roughly 1.30pm local time on May 1-16, 2003. Data are considered daily ($t = 1, \dots, 16$)
- **Global** coverage: Between -60° and 90° latitude. (Retrievals are not available below 60° S)
- $n_1 = 12515, n_2 = 12959, n_3 = 12971, n_4 = 12495, n_5 = 11862, n_6 = 12317,$
 $n_7 = 12577, n_8 = 12527, n_9 = 12651, n_{10} = 12252, n_{11} = 12681, n_{12} = 12076,$
 $n_{13} = 12245, n_{14} = 12447, n_{15} = 12372, n_{16} = 12532$
- We map our estimates onto a (hexagonal) grid of size $m_t = 57,065$ for each day t
- The measurement unit is parts per million (ppm)

AIRS CO_2 Data, ctd

Mid-tropospheric CO_2 on May 1-4 (for example), 2003, as measured by AIRS (gridded)



Assumptions and Specifications

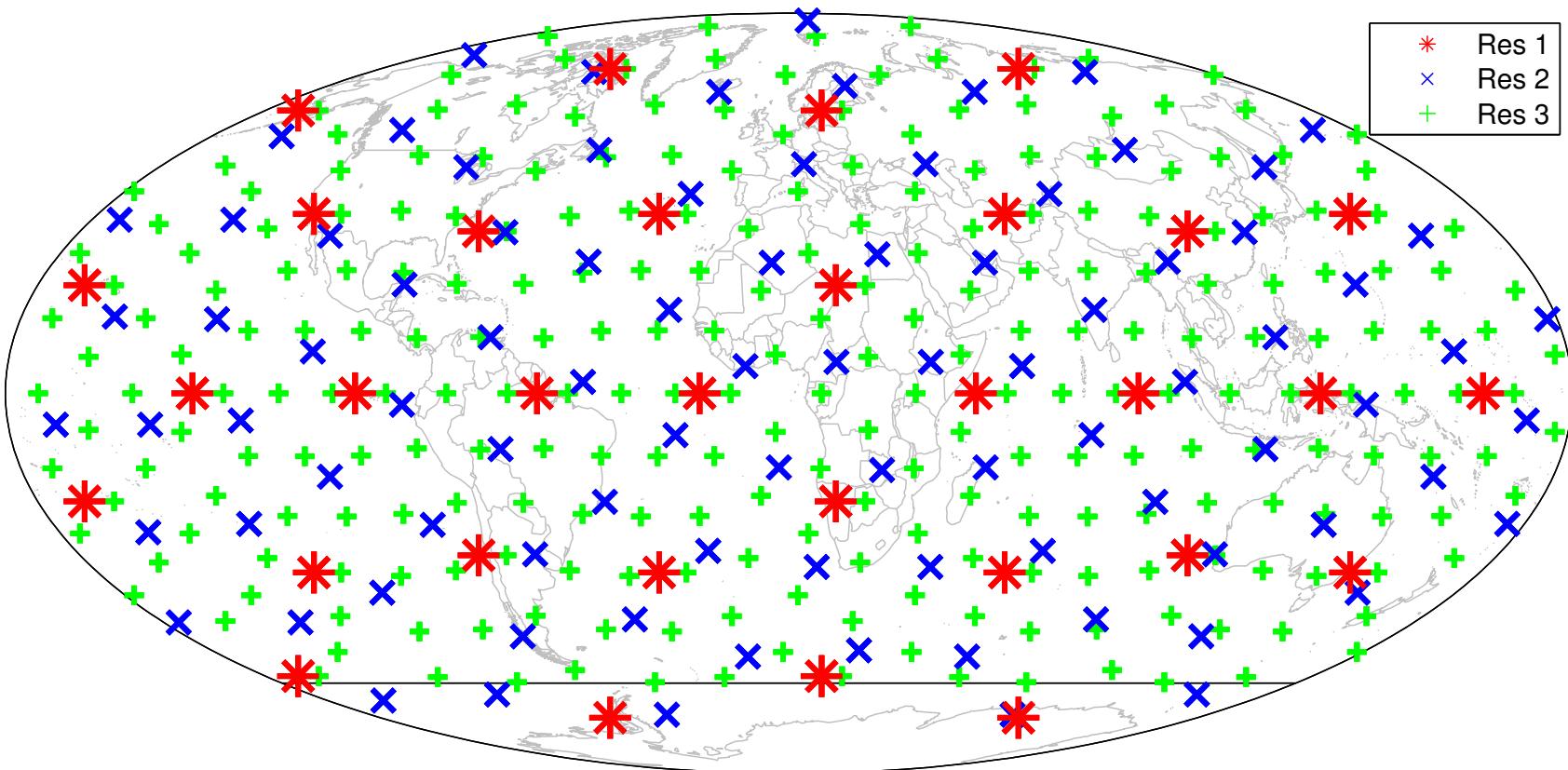
- Level 2 data were moved onto a fine regular hexagonal grid of approximately the resolution of the AIRS footprint. On some occasions, more than one Level 2 datum was in a grid cell: The datum $Z_t(\mathbf{s}_i)$ was obtained by averaging all Level 2 data in grid cell i on day t .
- Variances of measurement error and fine-scale variation:
 - $\sigma_\varepsilon^2 \equiv \sigma_{\varepsilon,1}^2 = \dots = \sigma_{\varepsilon,16}^2 = 5.6$ (estimated from the variogram)
 - $\sigma_\xi^2 \equiv \sigma_{\xi,1}^2 = \dots = \sigma_{\xi,16}^2$
 - $v_{\varepsilon,t}(\mathbf{s}_i)$ is equal to the inverse of the number of Level 2 data that were averaged over the i -th (hexagonal) grid cell on day t
- Basis functions: $r = 380$ bisquare functions at three spatial resolutions, the same for all $t = 1, \dots, 16$. The core basis function is the bisquare function:

$$b(\mathbf{u}) = (1 - \|\mathbf{u} - \mathbf{s}\|^2)^2 I(\|\mathbf{u} - \mathbf{s}\| \leq 1)$$

- Trend: $\mathbf{x}(\mathbf{s}) = (1, \text{lat}(\mathbf{s}))'$
- Make maps on a hexagonal grid of size $m_t = 57,065$ for each day $t = 1, \dots, 16$

Basis Functions: Bisquare Centers

Basis function centers, for all three resolutions



EM-FRS Results

- 21 iterations until convergence
- Approx. 4 min. per iteration on an eight-core server
- Total computing time for EM estimation: **96 min.** for 16 days of data
- Total computing time to obtain estimates of $\{Y_t(\cdot) : t = 1, \dots, 16\}$ using FRS: **15 min.** for 16 days of data

Overall computing time for **EM-FRS: 111 min.** for 16 days of data

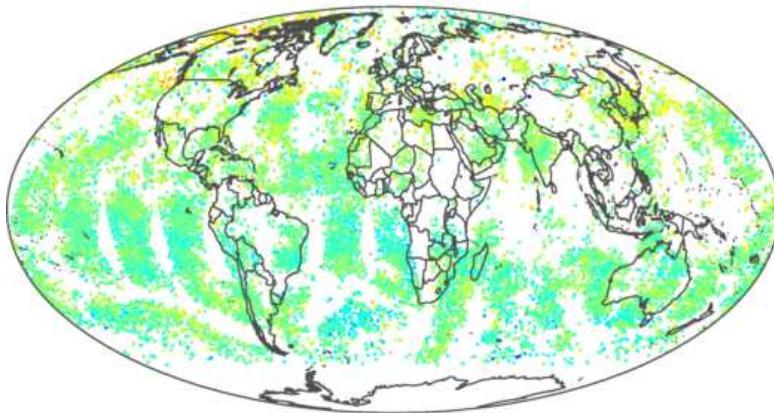
AIRS CO₂ animations are available:

www.stat.osu.edu/~sses/collab_co2.html

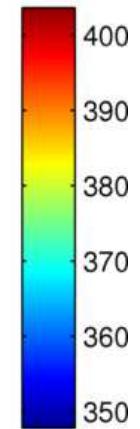
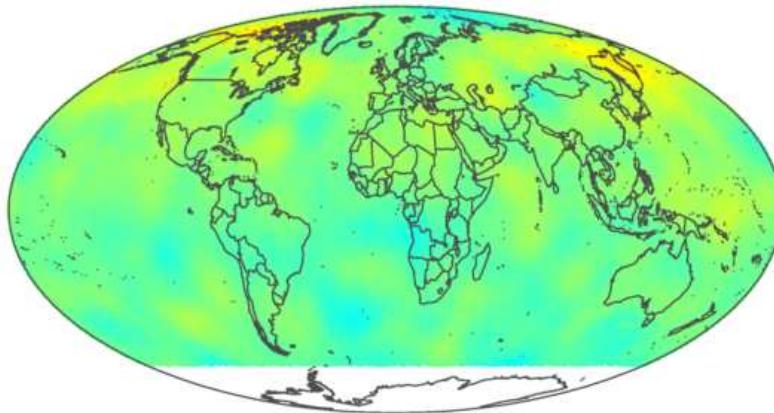
- The next slides show EM-FRS for May 1-16, 2003

EM-FRS Estimates, Day 1

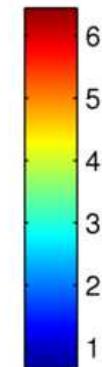
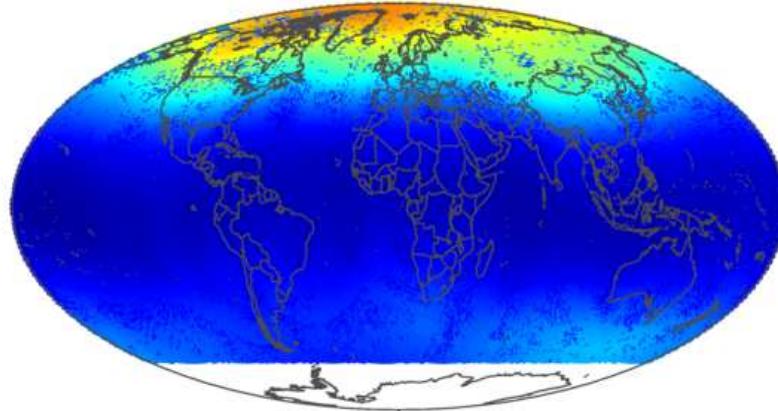
Gridded Data, day=1



EM Predictions, day=1

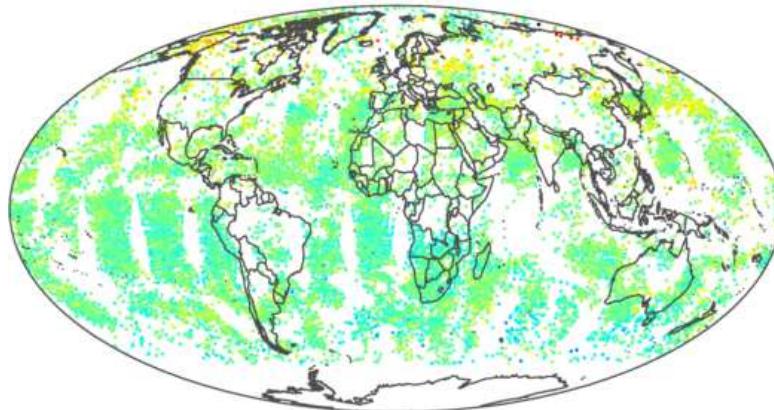


EM Standard Errors, day=1

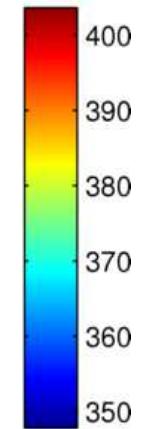
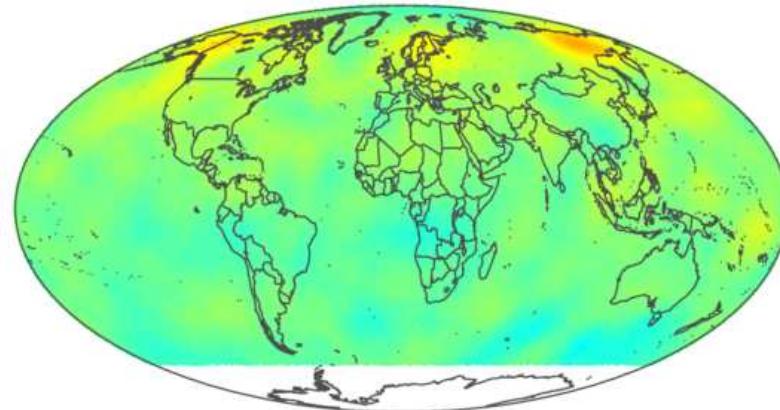


EM-FRS Estimates, Day 2

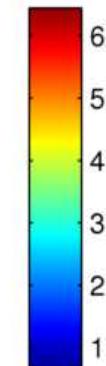
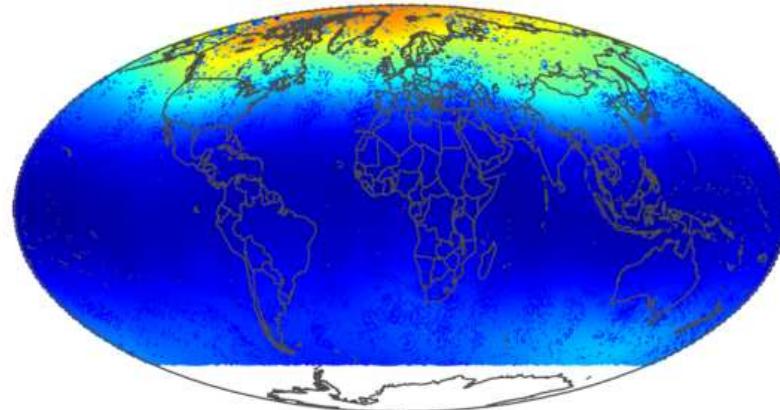
Gridded Data, day=2



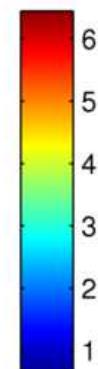
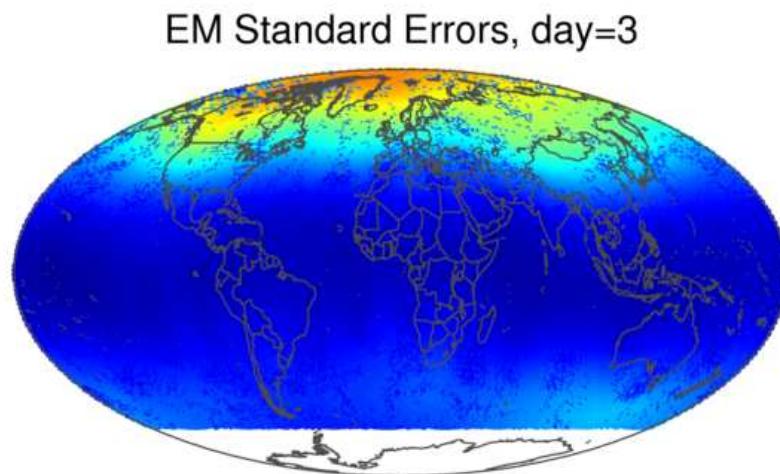
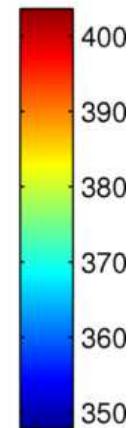
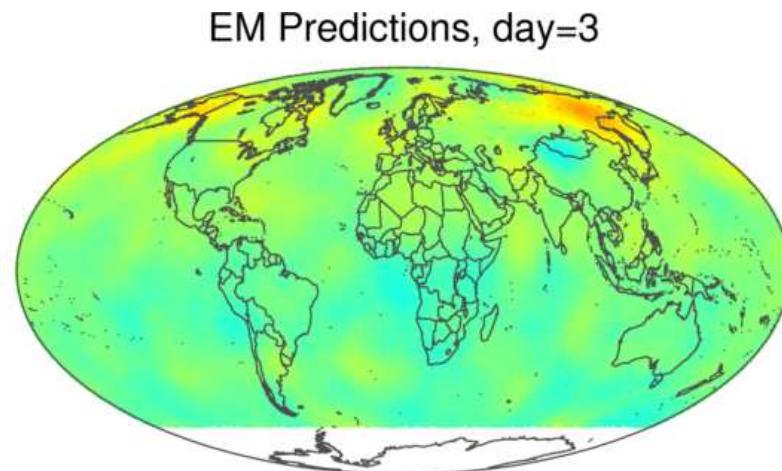
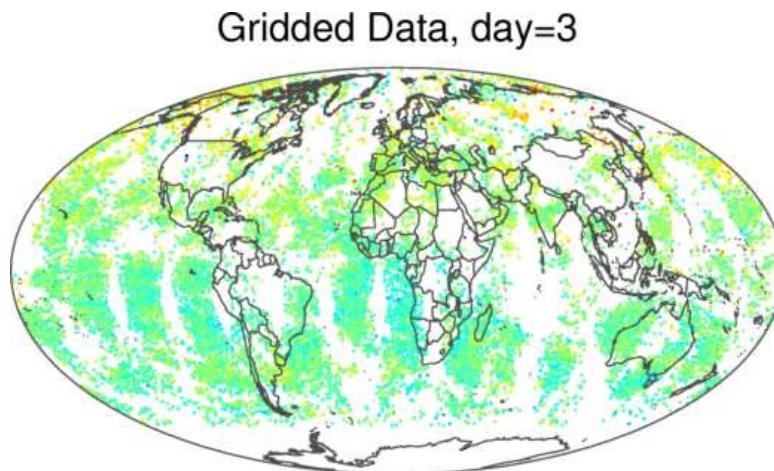
EM Predictions, day=2



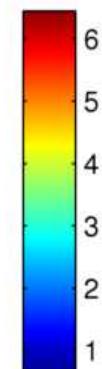
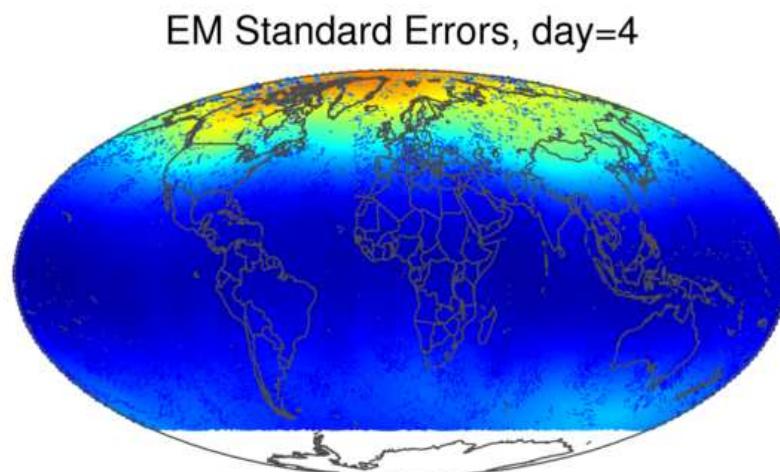
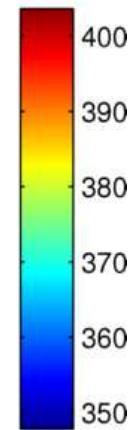
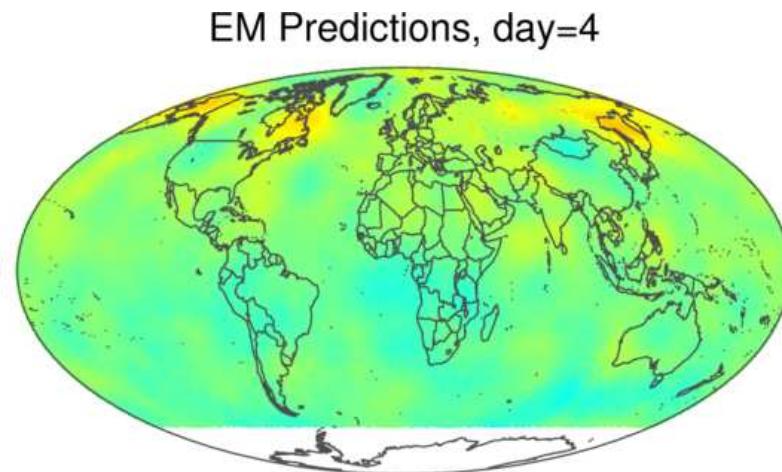
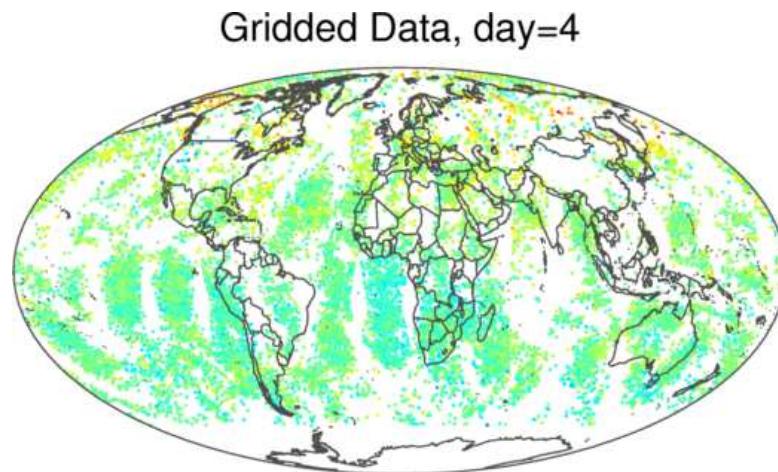
EM Standard Errors, day=2



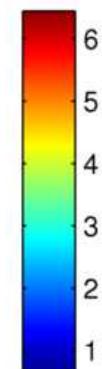
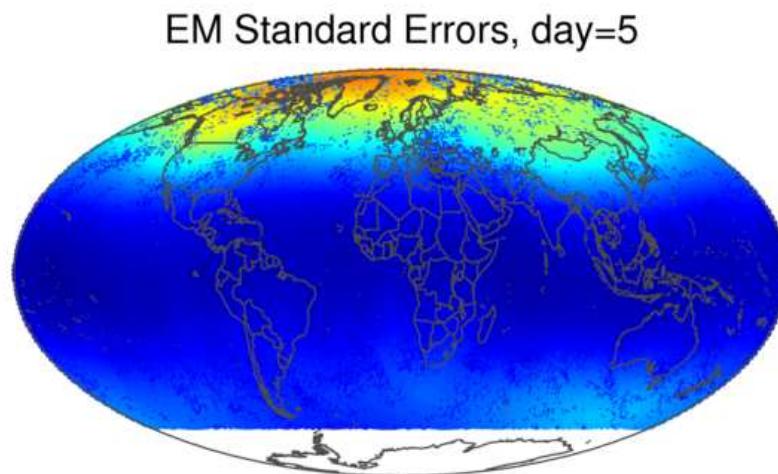
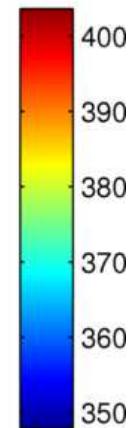
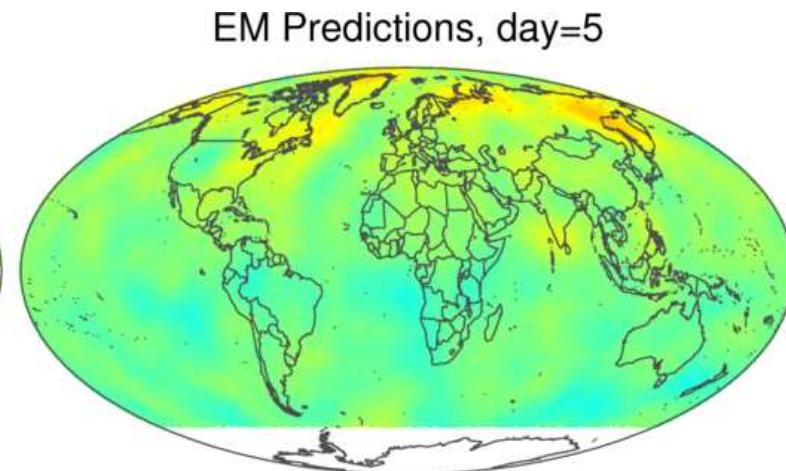
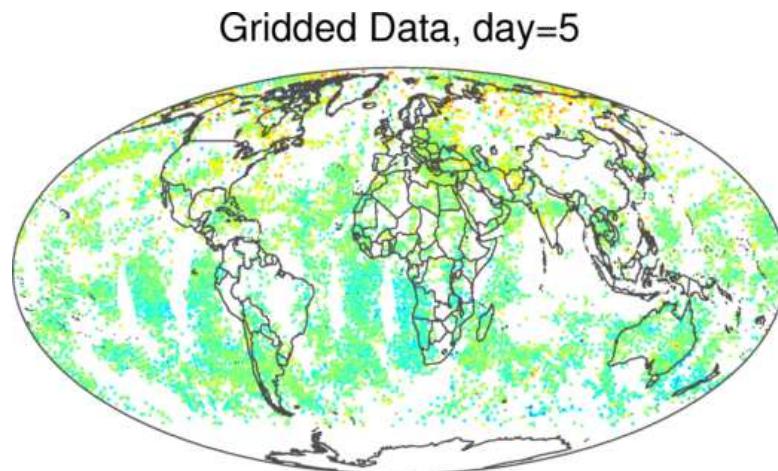
EM-FRS Estimates, Day 3



EM-FRS Estimates, Day 4

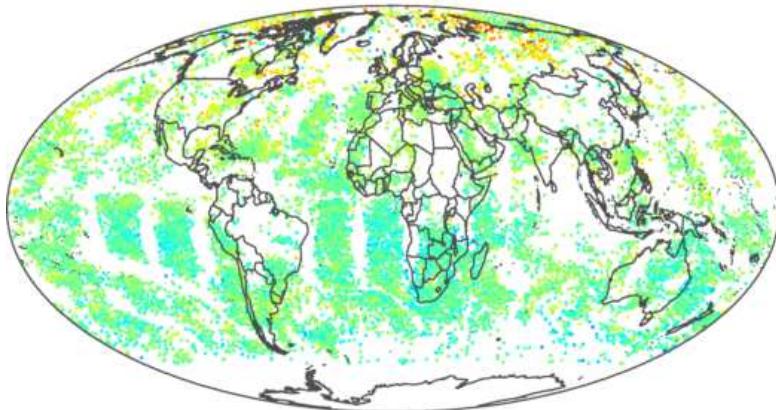


EM-FRS Estimates, Day 5

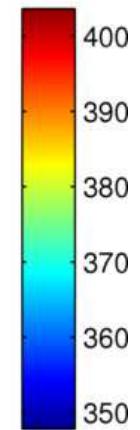
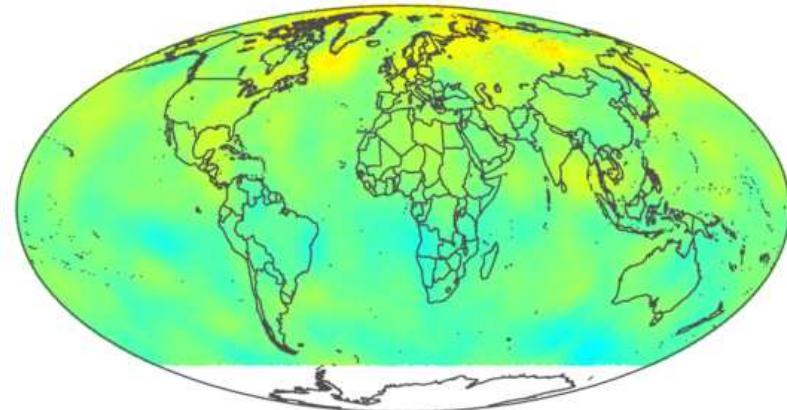


EM-FRS Estimates, Day 6

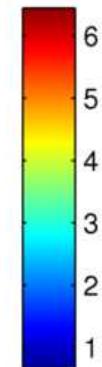
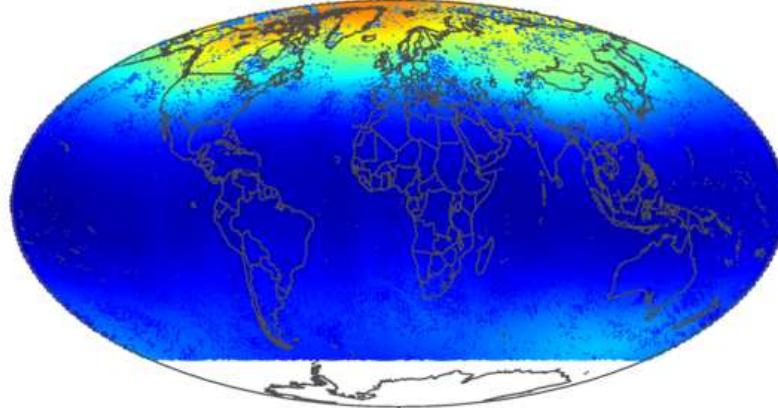
Gridded Data, day=6



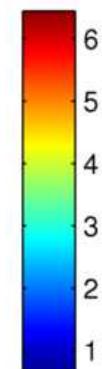
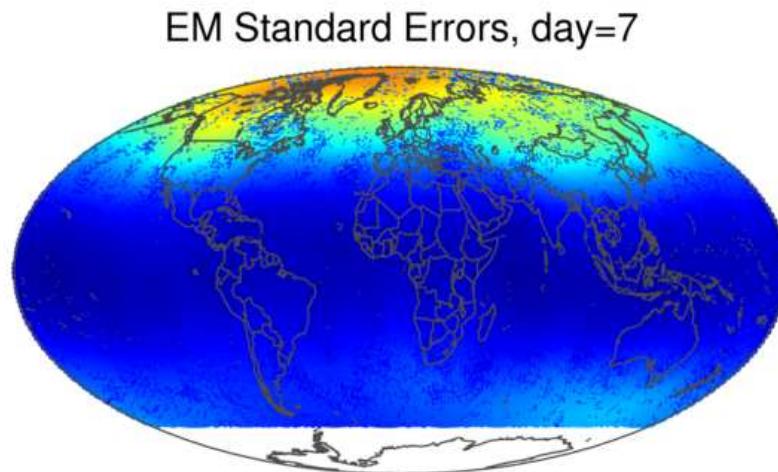
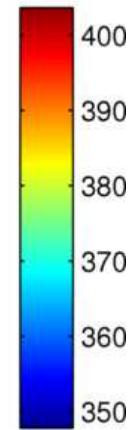
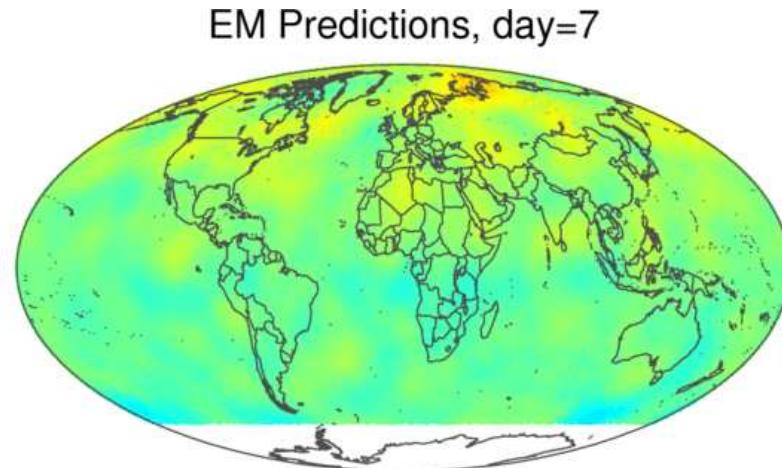
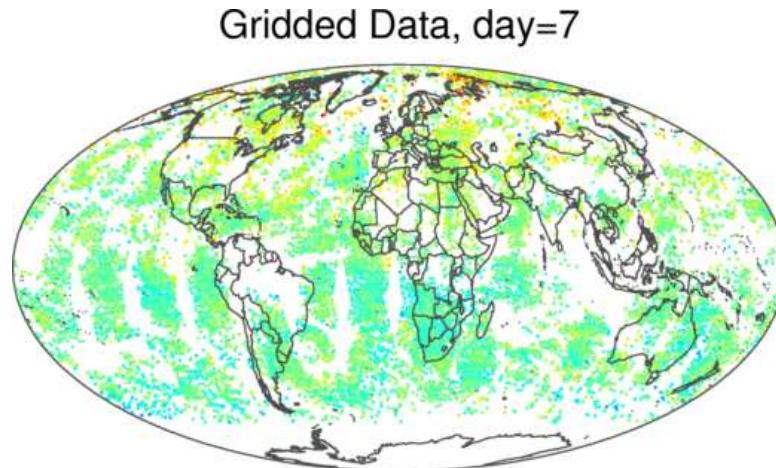
EM Predictions, day=6



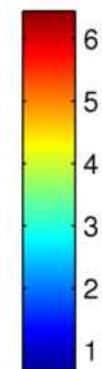
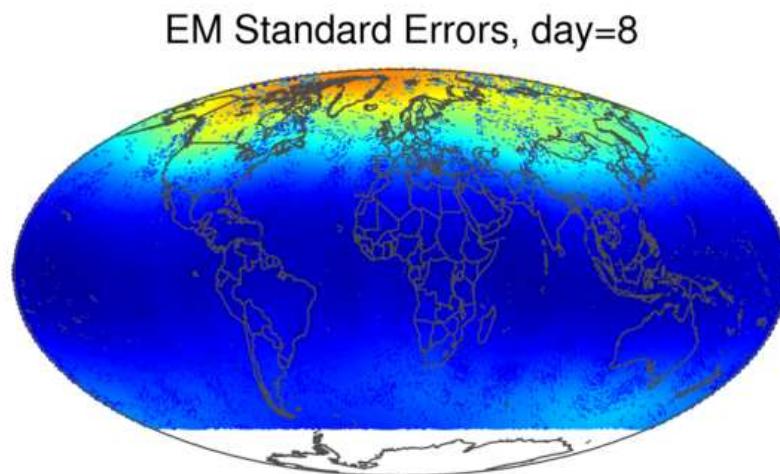
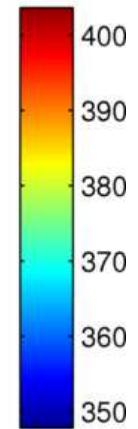
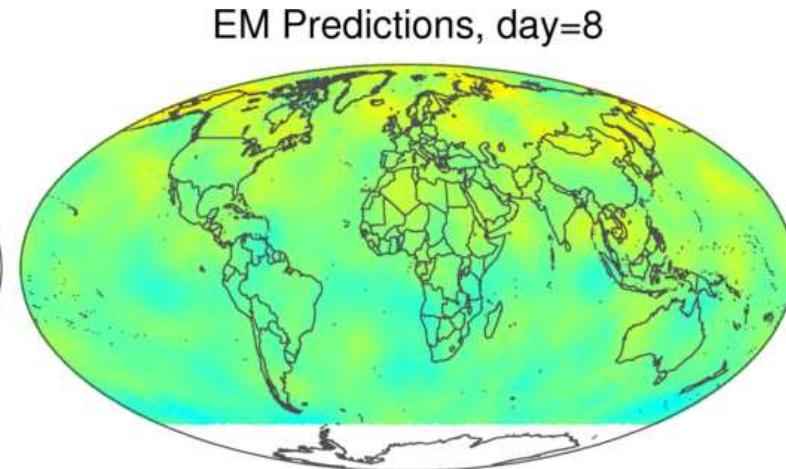
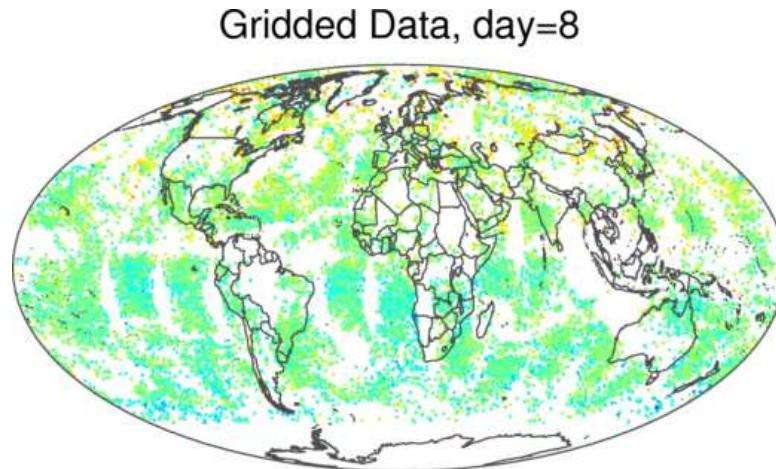
EM Standard Errors, day=6



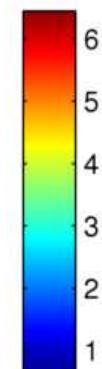
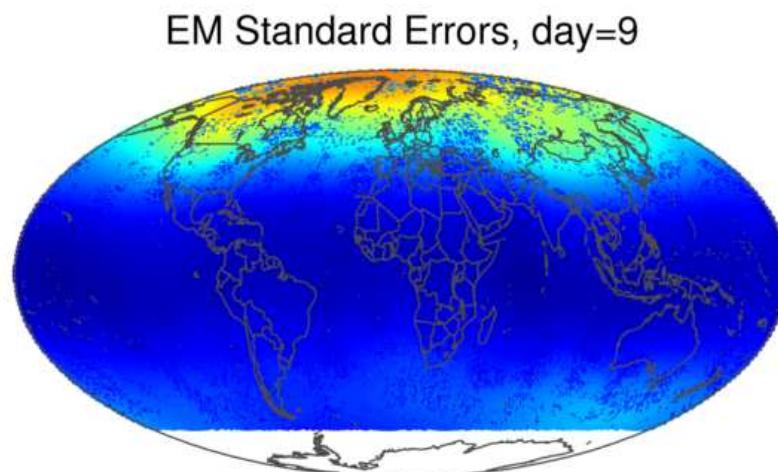
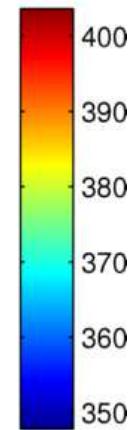
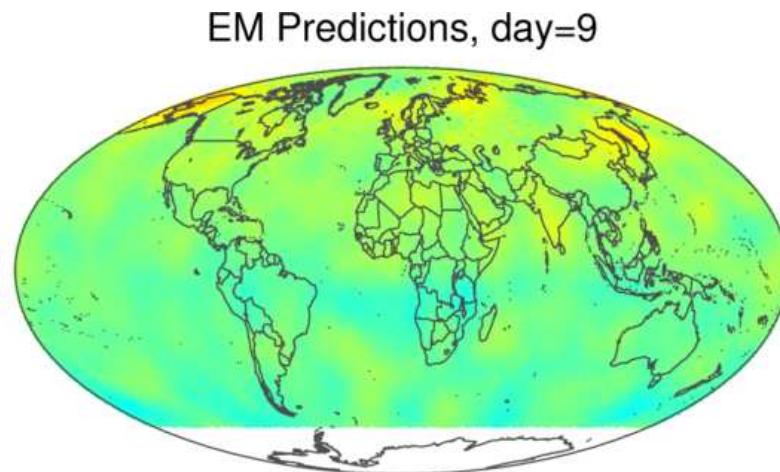
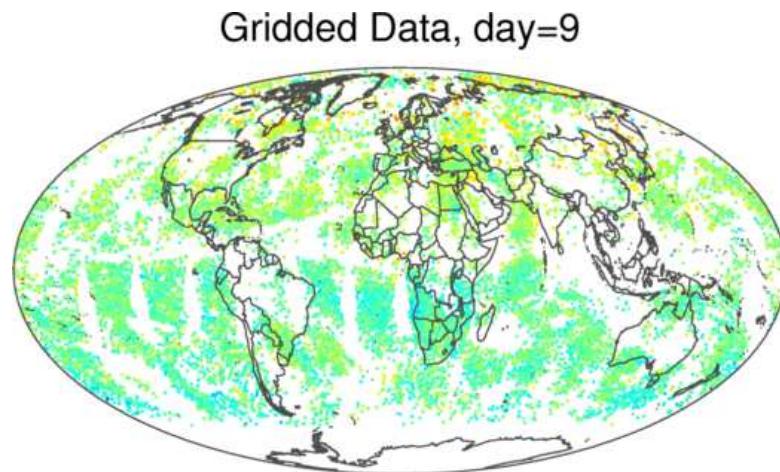
EM-FRS Estimates, Day 7



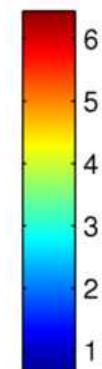
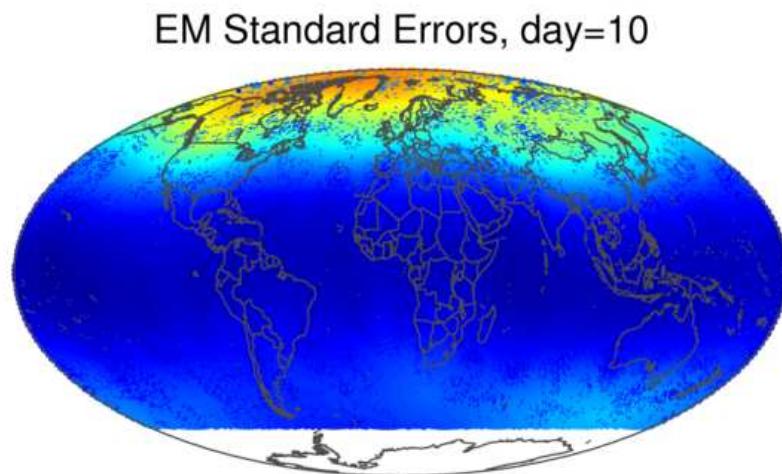
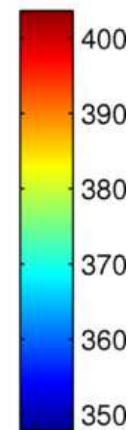
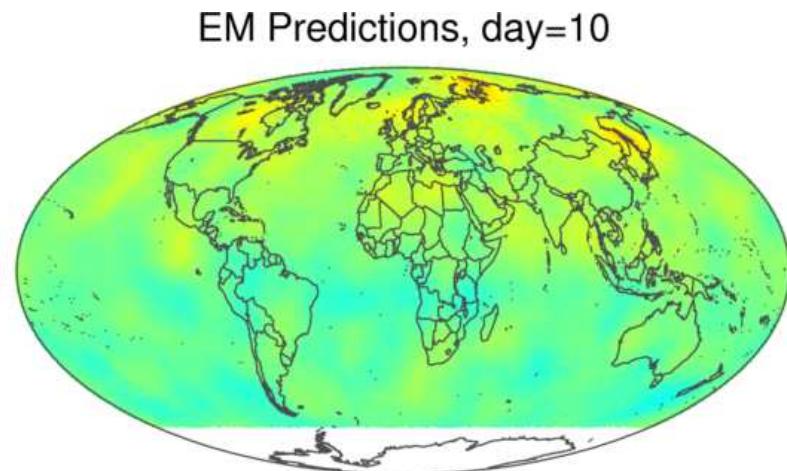
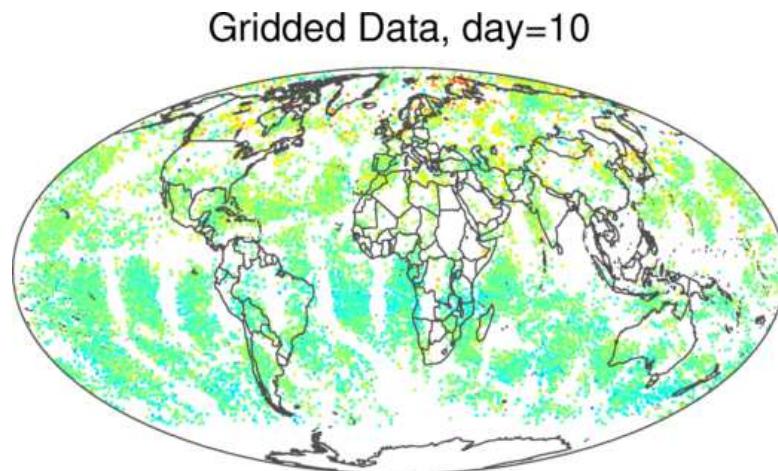
EM-FRS Estimates, Day 8



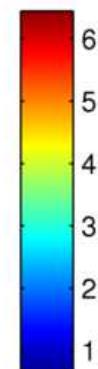
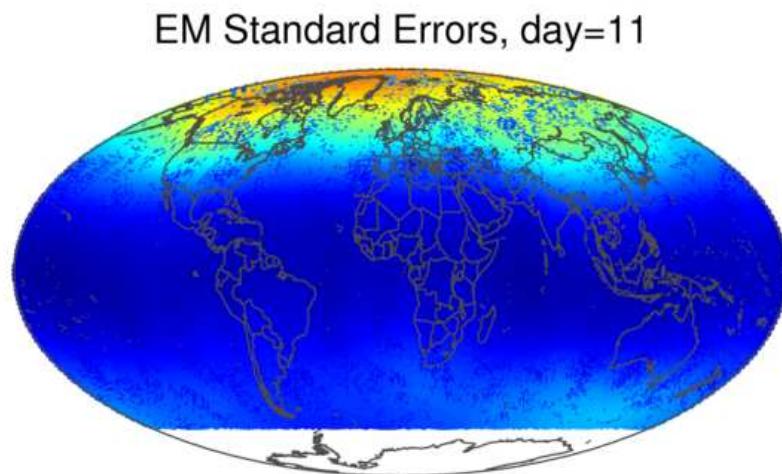
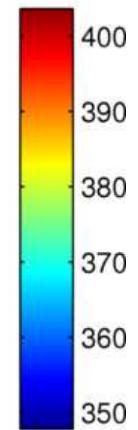
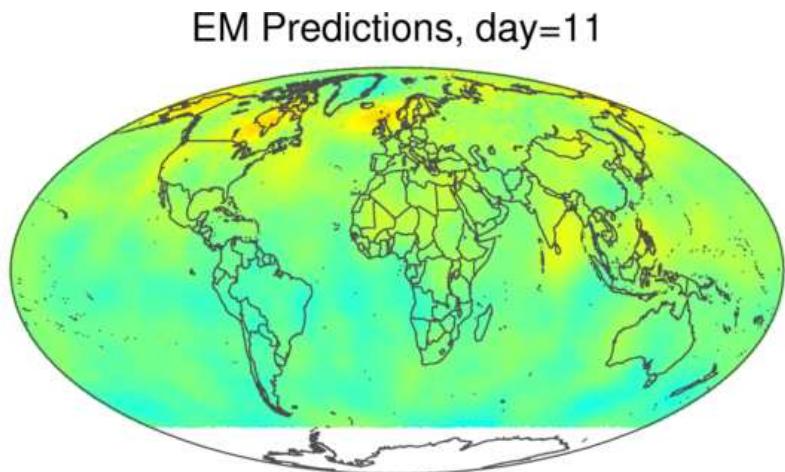
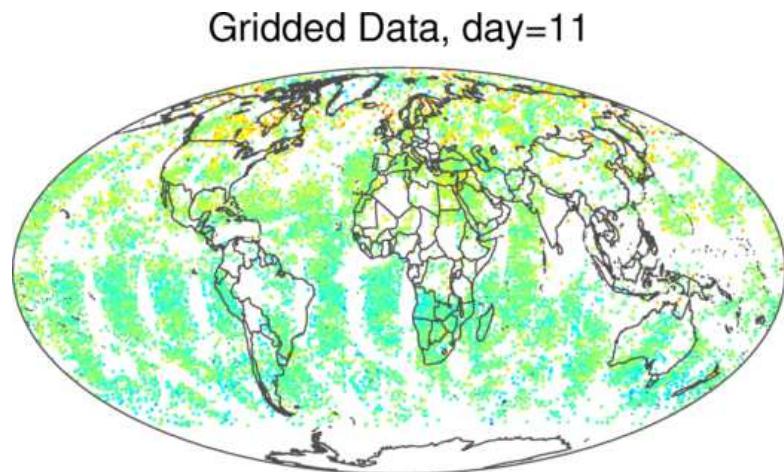
EM-FRS Estimates, Day 9



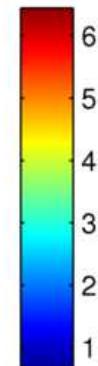
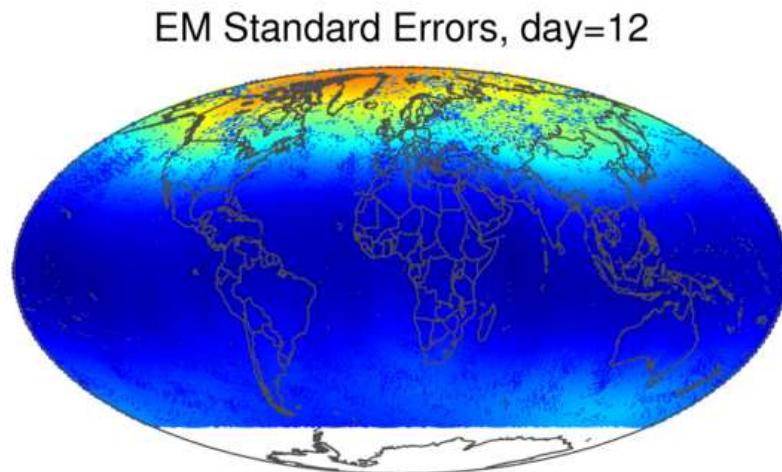
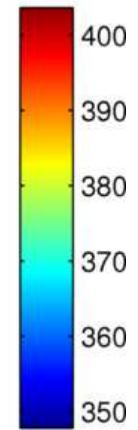
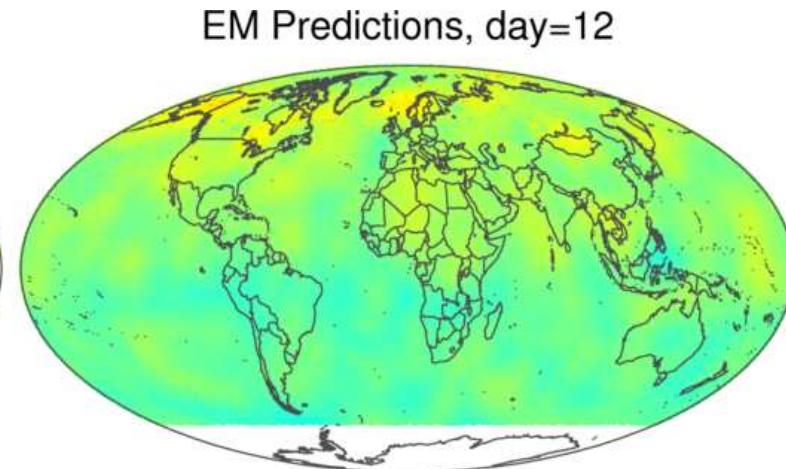
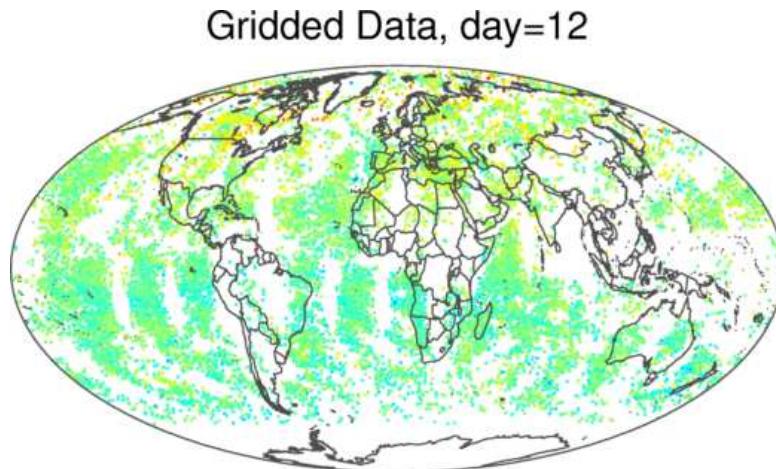
EM-FRS Estimates, Day 10



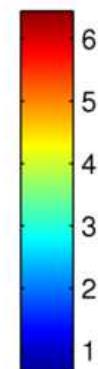
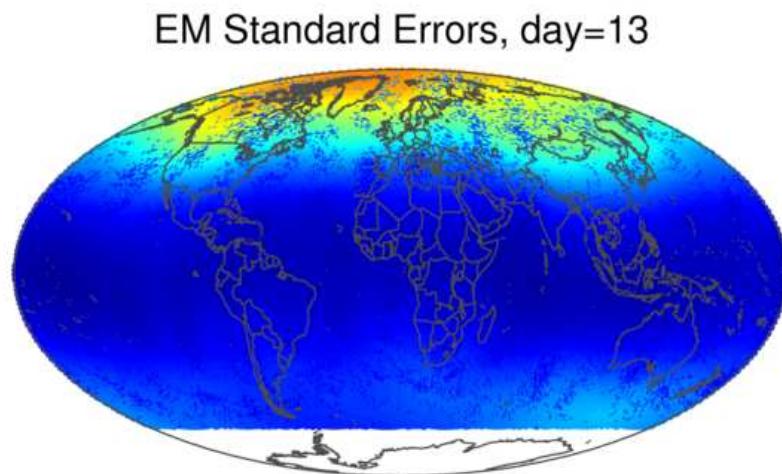
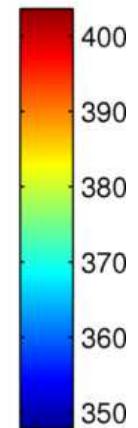
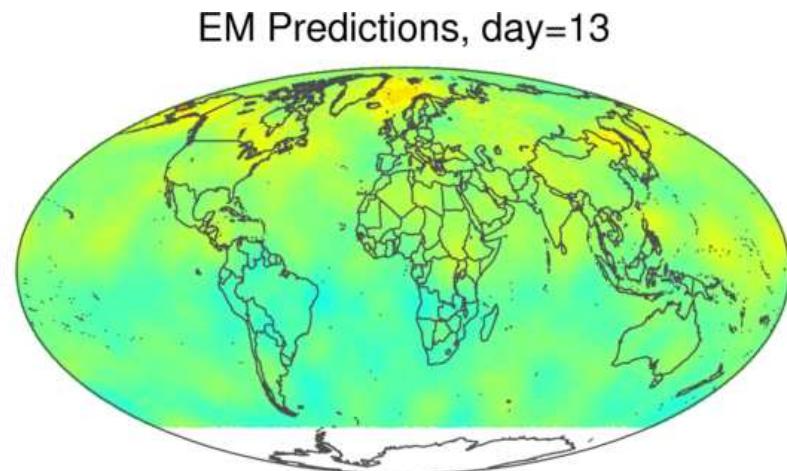
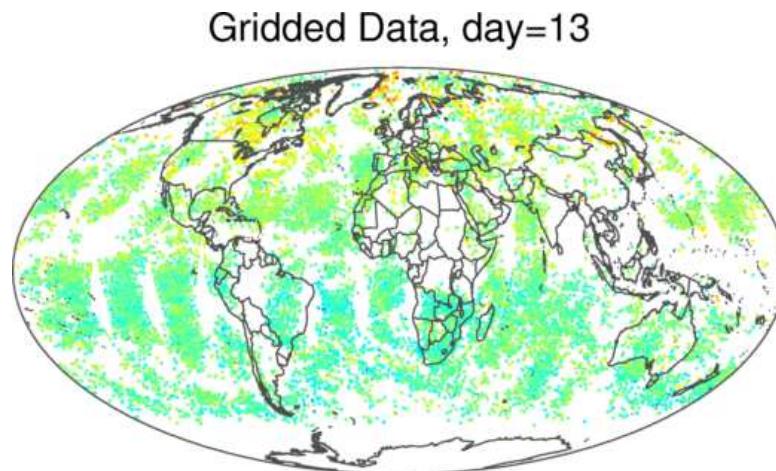
EM-FRS Estimates, Day 11



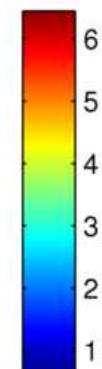
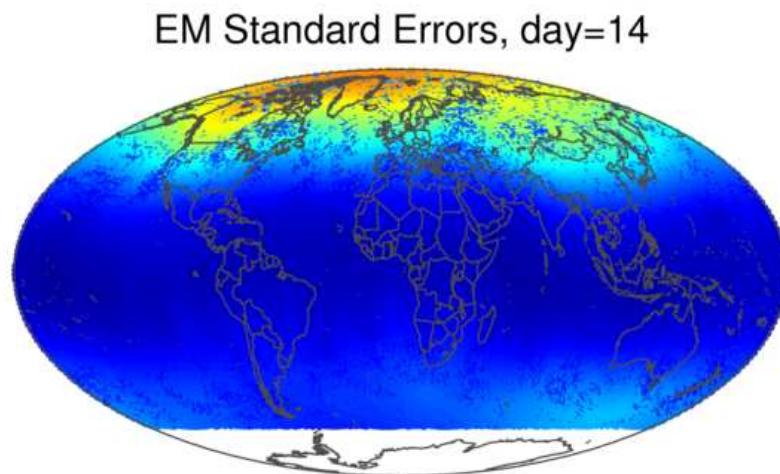
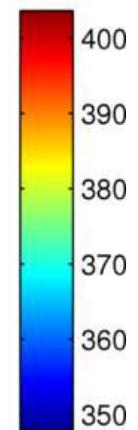
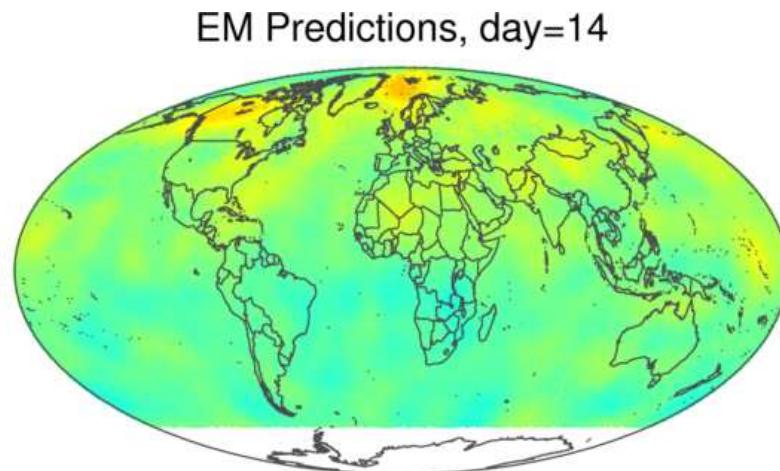
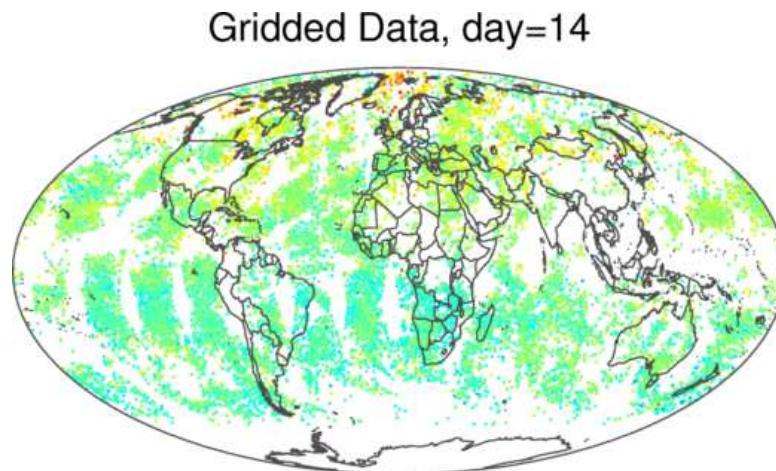
EM-FRS Estimates, Day 12



EM-FRS Estimates, Day 13

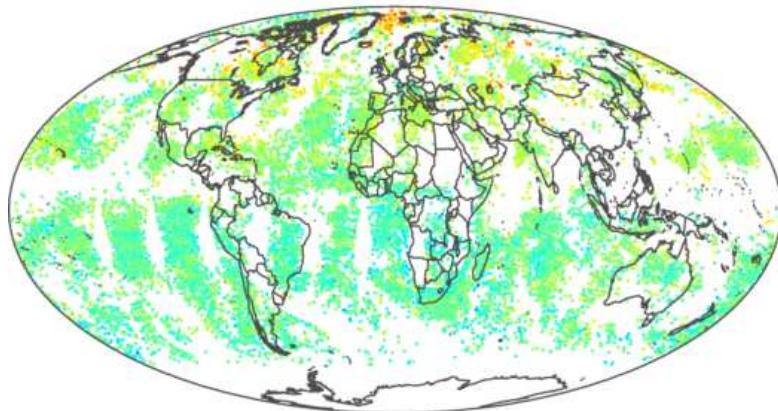


EM-FRS Estimates, Day 14

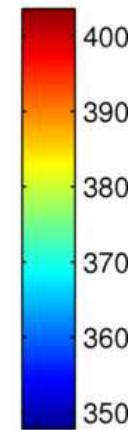
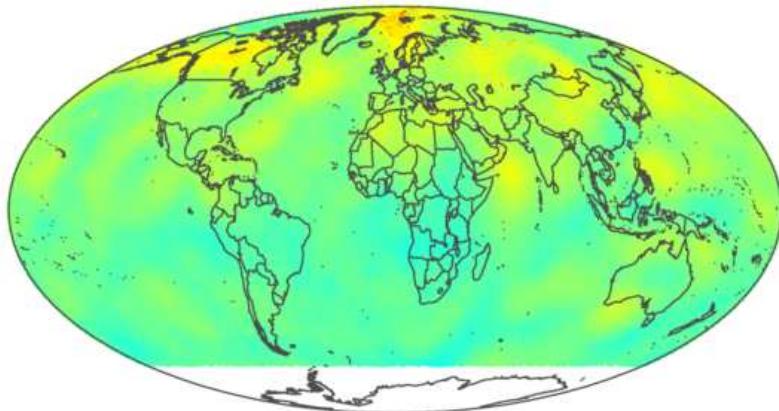


EM-FRS Estimates, Day 15

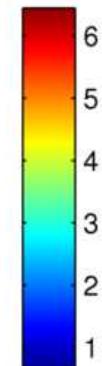
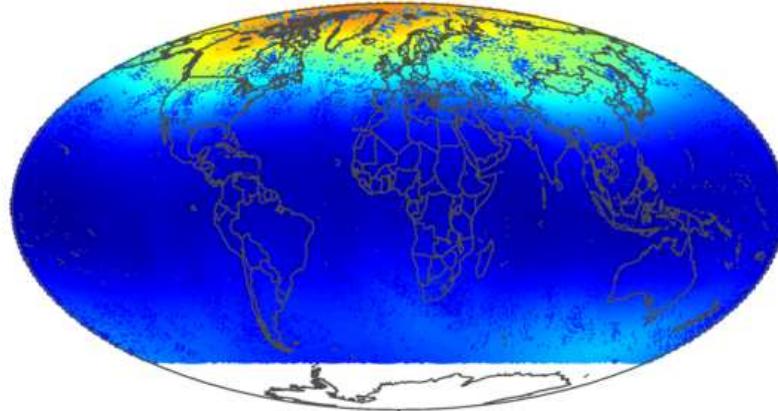
Gridded Data, day=15



EM Predictions, day=15

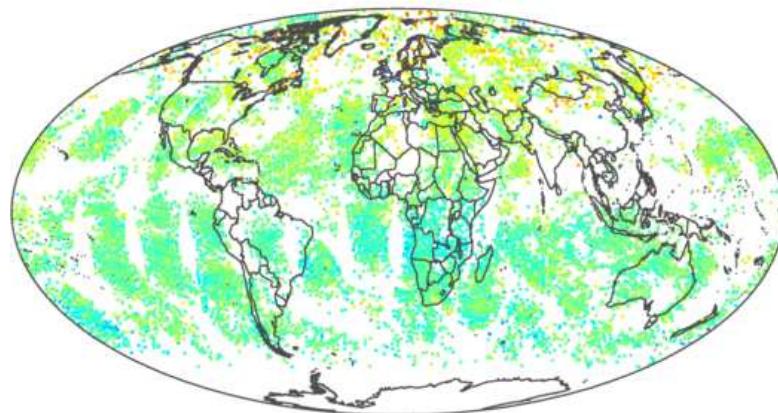


EM Standard Errors, day=15

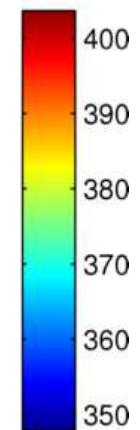
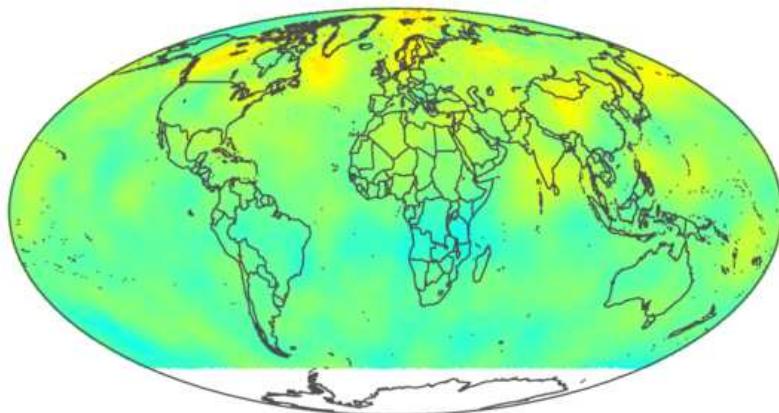


EM-FRS Estimates, Day 16

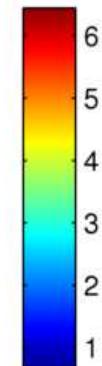
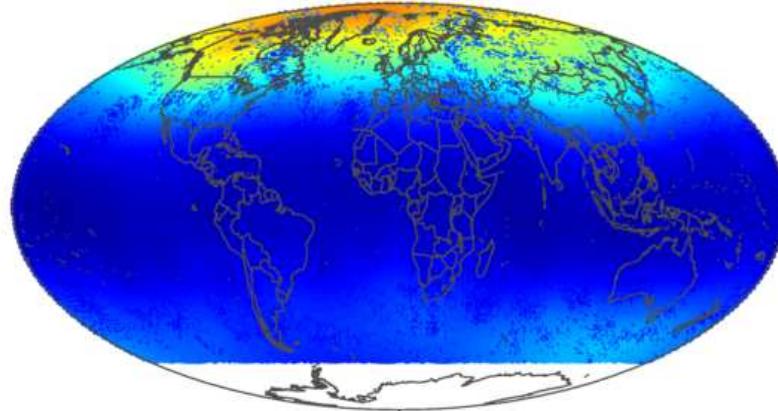
Gridded Data, day=16



EM Predictions, day=16



EM Standard Errors, day=16



Conclusions

The **spatio-temporal random effects (STRE)** model and **Fixed Rank Smoothing (FRS)**:

- handles massive datasets
- allows rapid computation of estimates
- allows for nonstationarity
- provides estimates of uncertainty, namely $(MSPE)^{1/2}$

Parameter estimation

- **EM** allows rapid computation and requires minimal user input

Further details:

- Noel Cressie (ncressie@stat.osu.edu)
- An SSES Web-Project: www.stat.osu.edu/~sses/collab_co2.html

References

- Cressie, N. (1993). *Statistics for Spatial Data*, rev. edn. Wiley, New York, NY.
- Cressie, N., Shi, T., and Kang, E.L. (2010). *Journal of Computational and Graphical Statistics*, **19**, 724-745.
- Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.
- Katzfuss, M. and Cressie, N. (2009). In *2009 Proceedings of the Joint Statistical Meetings*. American Statistical Association, Alexandria, VA, 3378-3390.
- Katzfuss, M. and Cressie, N. (2011a). *Journal of Time Series Analysis*, **32**, 430-446.
- Katzfuss, M. and Cressie, N. (2011b). OSU Dept. of Statistics Tech Report No. **853**. Under revision for *Environmetrics*.
- Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications, with R Examples*, 2nd edn. Springer, New York, NY.
- Wu, C.F.J. (1983). *Annals of Statistics*, **11**, 95-103.